Fast robustness quantification with variational Bayes

Ryan Giordano, Tamara Broderick, Rachael Meager, Jonathan Huggins, Michael Jordan

Presentation Outline

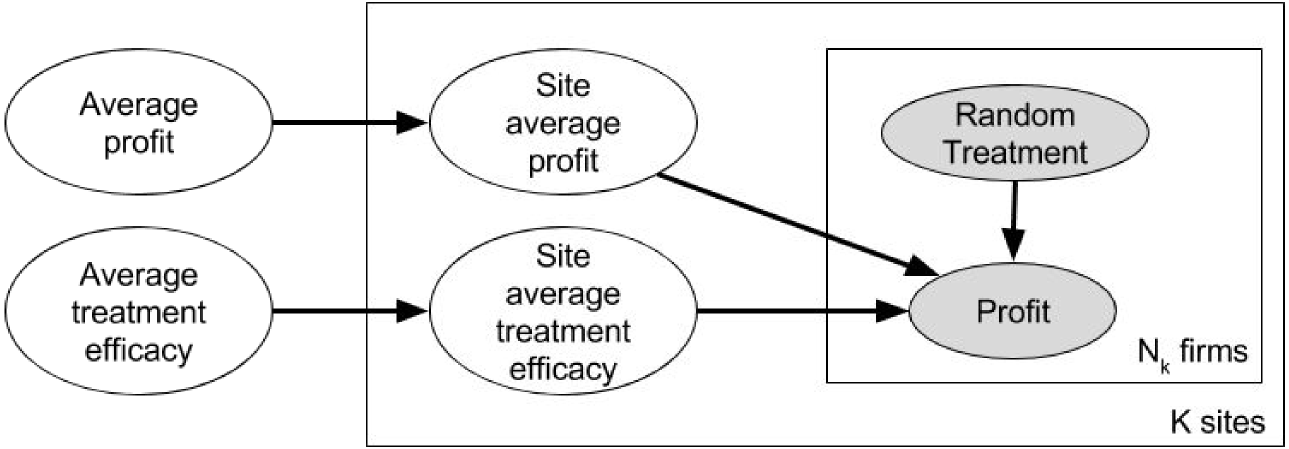
- Hierarchical models and motivating example
- Measuring robustness
- Variational Bayes
- Experiments
- Conclusions

Hierarchical Models

Randomized controlled trials were conducted in different locations.

Key idea: pool the data with a Bayesian hierarchical model.

- Seven sites
- ~37,000 total observations
- From ~1,000 to ~17,000 observations per site

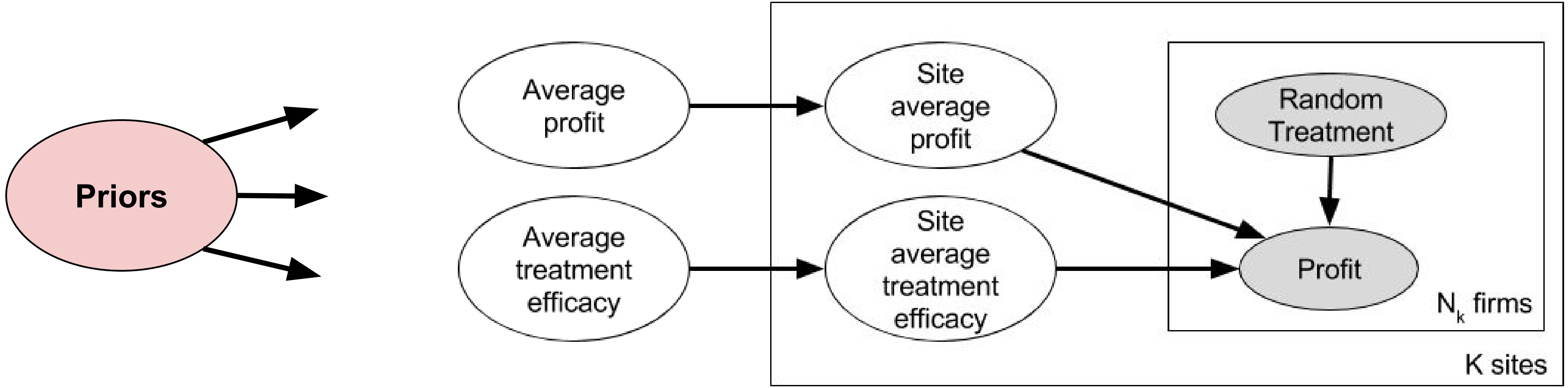


Understanding the impact of microcredit expansions: A Bayesian hierarchical analysis of 7 randomised experiments. Rachael Meager (2015)

Hierarchical Models

Key question:

Bayesian models require priors. How robust is our analysis to the choice of priors?



Understanding the impact of microcredit expansions: A Bayesian hierarchical analysis of 7 randomised experiments. Rachael Meager (2015)

Hierarchical Models

$$\begin{aligned}
\sigma_k^{-2} &\sim \mathrm{Gamma}\left(\alpha_\tau, \beta_\tau\right) \\
C &\sim \mathrm{LKJ}\left(\eta, \alpha, \gamma\right) \\
\begin{pmatrix} \mu \\ \tau \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} \mu_0 \\ \tau_0 \end{pmatrix}, \Lambda^{-1}\right)
\end{aligned}$$

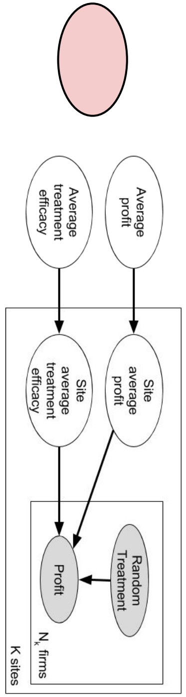
Priors

$$\begin{pmatrix} \mu_k \\ \tau_k \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \tau \end{pmatrix}, C\right)$$

Hierarchy

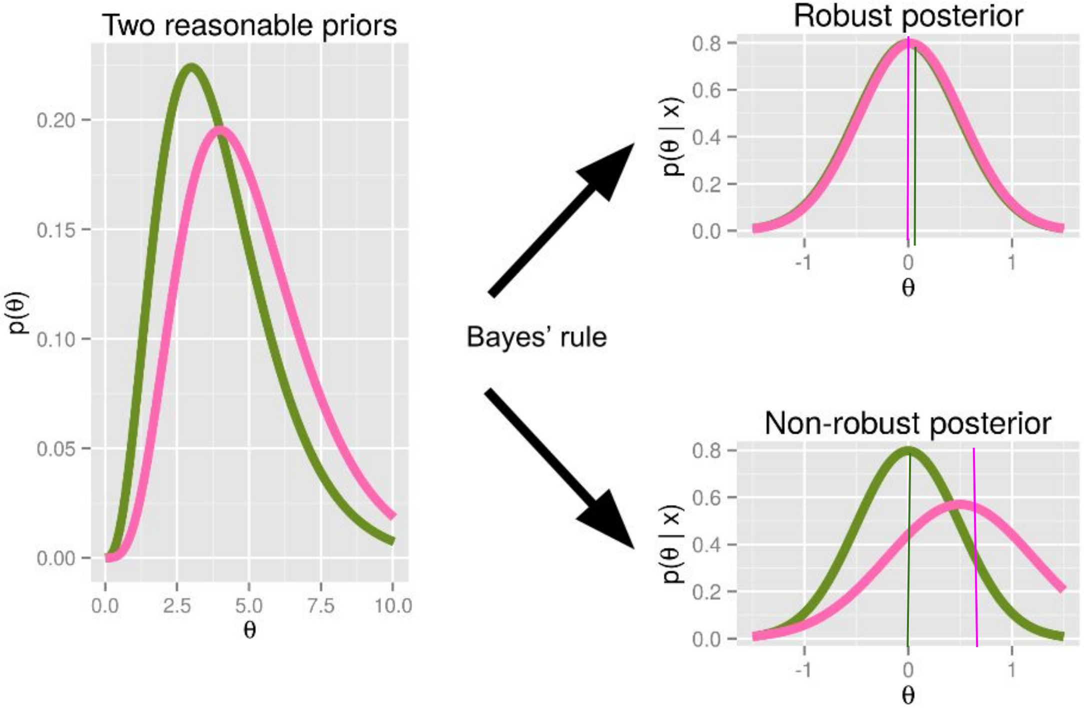
$$y_{nk} \sim \mathcal{N}\left(\mu_k + T_{ik}\tau_k, \sigma_k^2\right)$$

Observations



Understanding the impact of microcredit expansions: A Bayesian hierarchical analysis of 7 randomised experiments. Rachael Meager (2015)

How do we measure robustness?



Local sensitivity:

Posterior expectation of interest

$$\frac{d\mathbb{E}_{p_x^\alpha}\left[\theta\right]}{d\alpha}$$

Prior parameters
(vector or function-valued)

Local robustness in Bayesian analysis (in **Robust Bayesian Analysis**)
Paul Gustafson (2012)

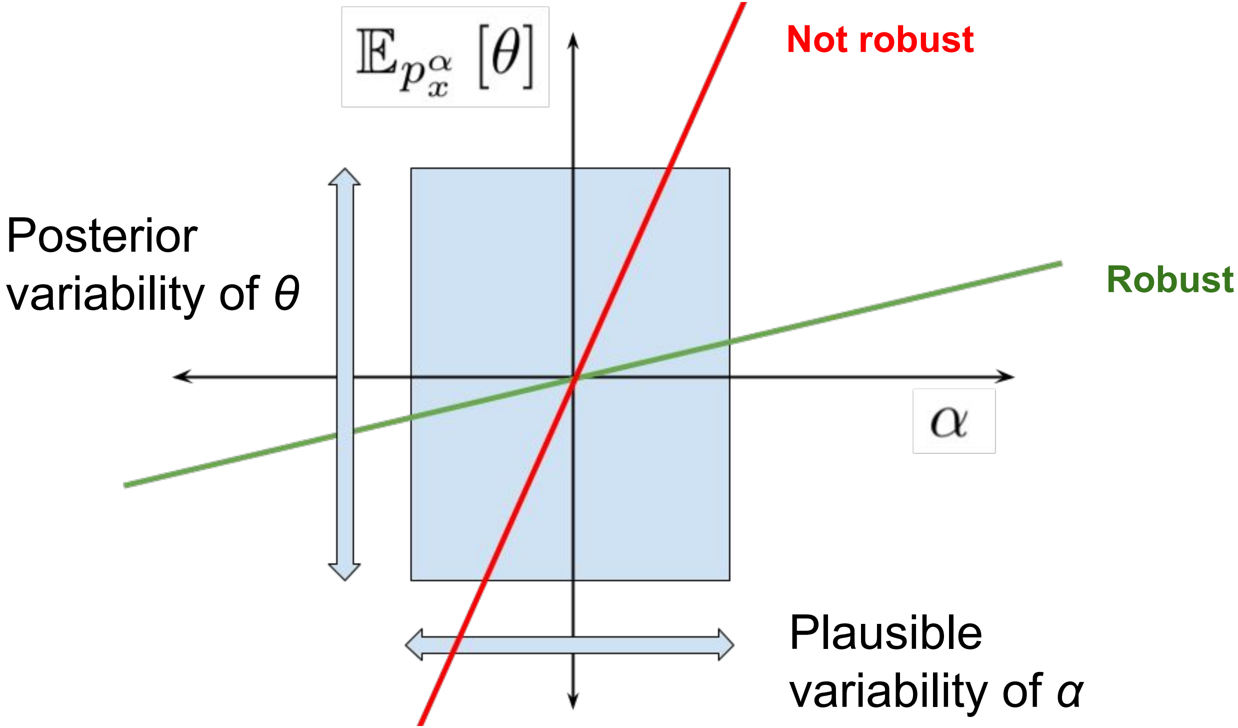
How do we use local sensitivity?

Local sensitivity:

Posterior expectation of interest

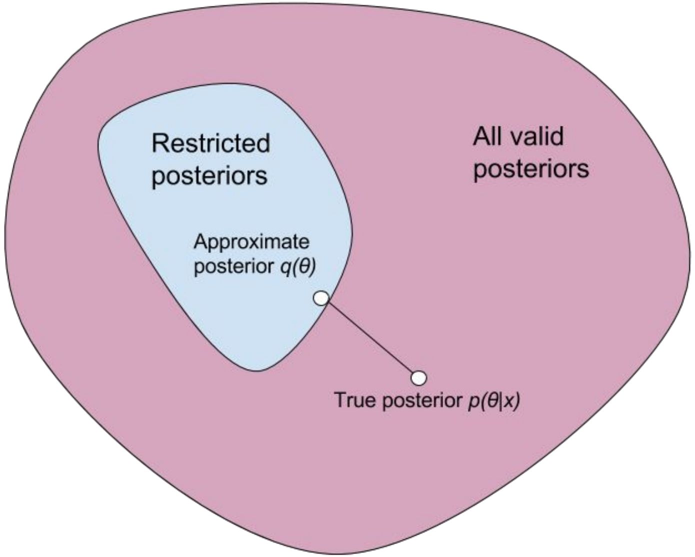
$$\frac{d\mathbb{E}_{p_x^\alpha}[\theta]}{d\alpha}$$

Prior parameters
(vector or function-valued)



The modeler still needs to decide how much α might vary.

What is variational Bayes?



We want $p(\theta|x)$ a posterior.
Find the closest distribution in a simpler family:

$$q(\theta) = \underset{q' \in \mathcal{Q}}{\operatorname{argmin}} \; KL\left(q' || p(\theta|x)\right)$$
$$\mathcal{Q} = \{\text{Some restricted class of distributions}\}$$

If distributions in the family factorize across variables, we call it "mean field variational Bayes" (MFVB)

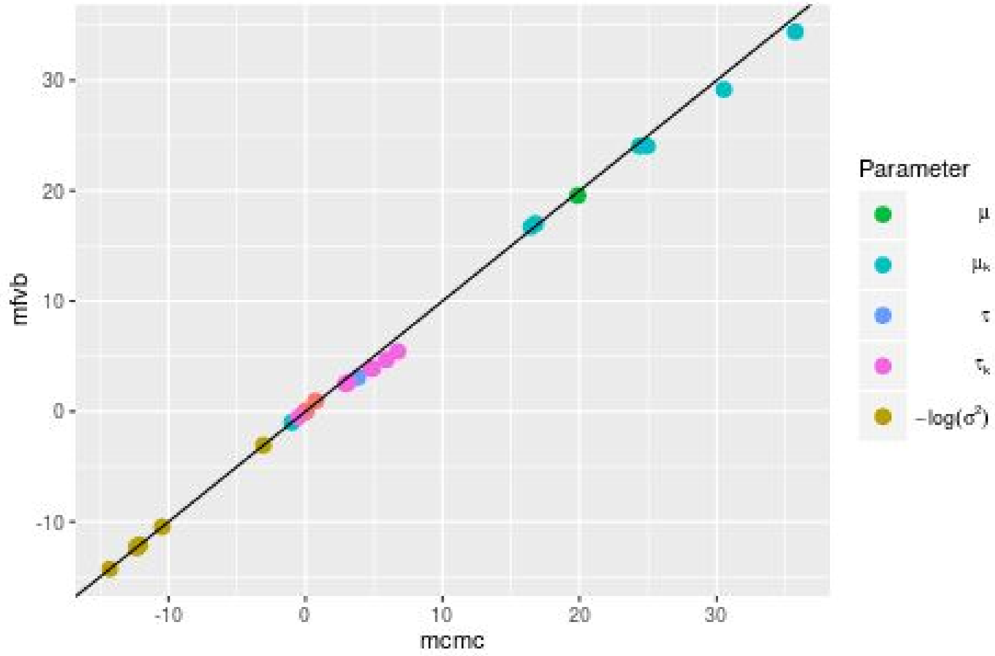
Variational Inference: A Review for Statisticians
David M Blei, Alp Kucukelbir, Jon D McAuliffe (2016)

Why use variational Bayes?

It's very fast. On our model:

```
MCMC time (with Stan): 45 minutes
VB time:                52 seconds
```

Often, the variational posterior means match MCMC quite closely.

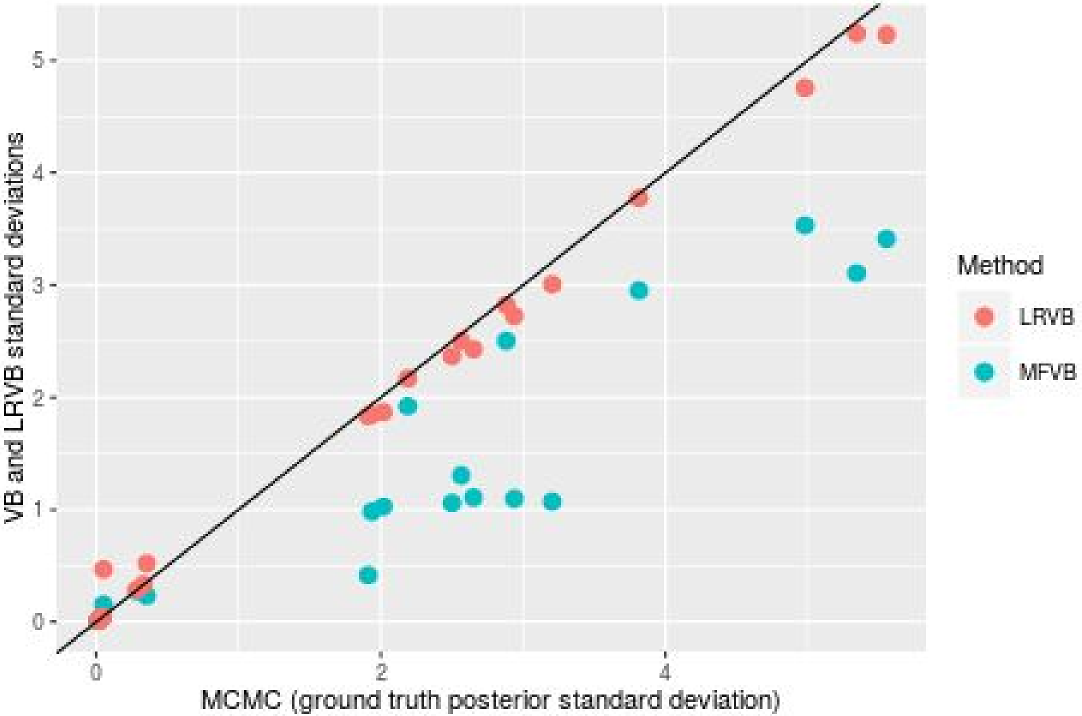


Why not use variational Bayes?

If MFVB is wrong, it's hard to know how wrong it is without running MCMC anyway.

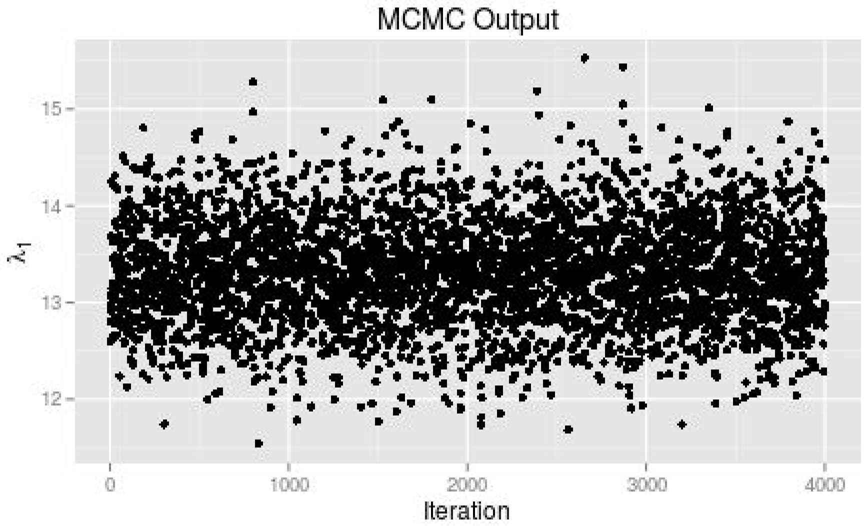
=> Mostly useful for prototyping.

Historically, MFVB was not used for inference because it tends to underestimate posterior variance. However, we have a fix.

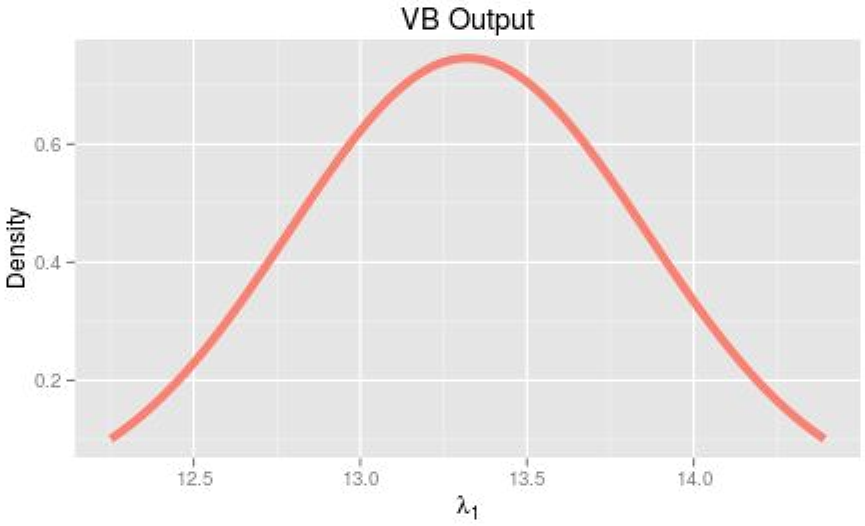


Linear response methods for accurate covariance estimates from mean field variational Bayes, Ryan Giordano, Tamara Broderick, Michael Jordan (2015)

Variational Bayes is local-robustness friendly



MCMC: designed for integration
(sensitivity hard to measure in general)



VB: designed for differentiation
($\exists$ closed form sensitivity measures)

Variational Bayes is local-robustness friendly

$$\ell(\alpha, m) \quad := \quad \mathbb{E}_{q_x^\alpha}\left[\log p(\theta|\alpha)\right]$$

Expected log prior (a function of the prior parameters and the variational distribution parameters)

Variational Bayes is local-robustness friendly

$$
\begin{aligned}
\ell\left(\alpha, m\right) &:= \mathbb{E}_{q_x^\alpha}\left[\log p\left(\theta|\alpha\right)\right] \\
q_t &:= \mathrm{argmin}_{q \in \mathcal{Q}}\left\{KL + \frac{\partial \ell}{\partial \alpha^T}\delta_\alpha t + O\left(t^2\right)\right\}
\end{aligned}
$$

Optimize KL divergence between the original model with a locally perturbed prior

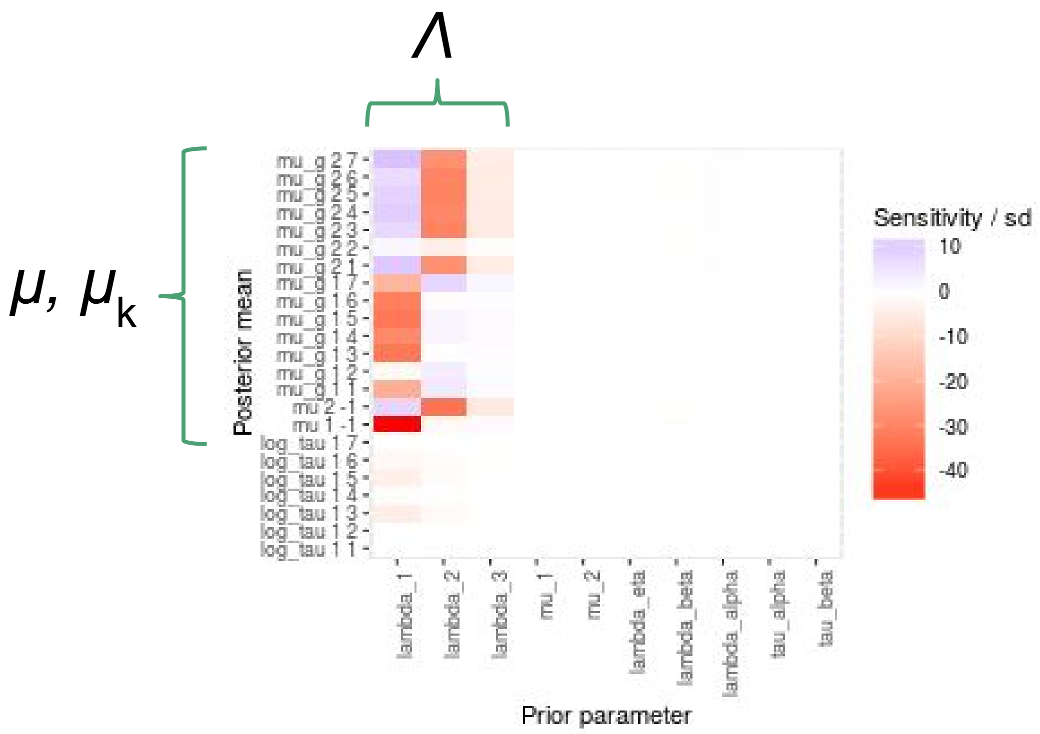
Variational Bayes is local-robustness friendly

$$
\begin{aligned}
\ell\left(\alpha, m\right) &:= \mathbb{E}_{q_x^\alpha}\left[\log p\left(\theta|\alpha\right)\right] \\
q_t &:= \mathrm{argmin}_{q\in\mathcal{Q}}\left\{KL + \frac{\partial \ell}{\partial \alpha^T}\delta_\alpha t + O\left(t^2\right)\right\} \\
\left.\frac{d\mathbb{E}_{q_x^\alpha}\left[\theta\right]}{dt}\right|_{t=0} &= \left(\frac{\partial^2 KL}{\partial m \partial m^T}\right)^{-1}\frac{\partial^2 \ell}{\partial m \partial \alpha^T}\delta_\alpha
\end{aligned}
$$

Local sensitivity is given by a linear system.

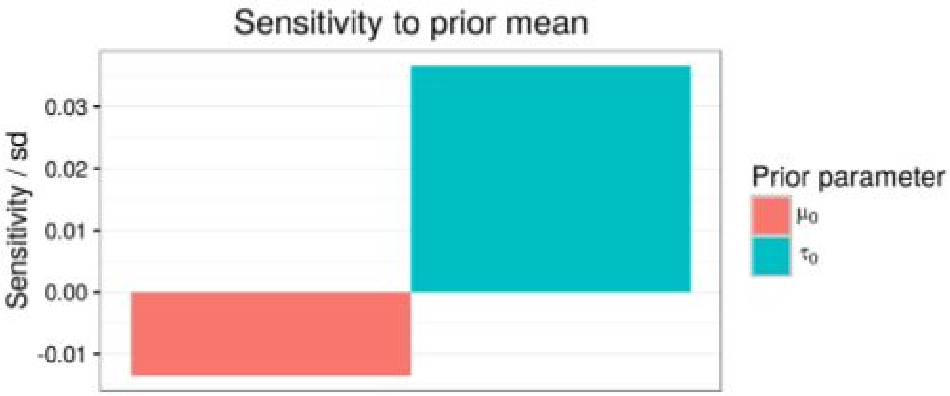
Experiments

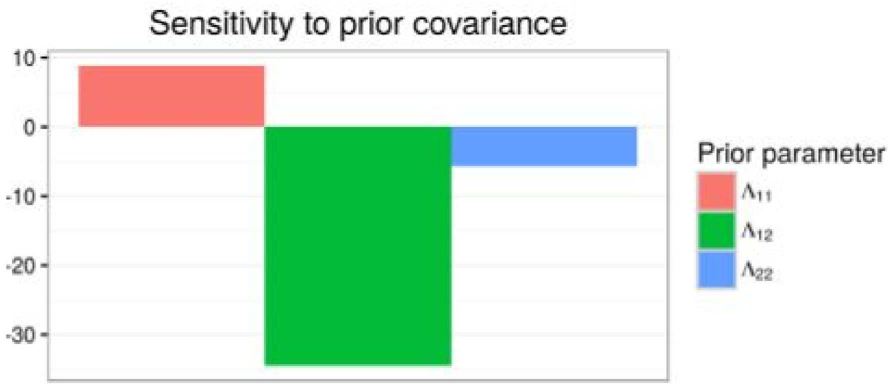
With our choice of priors, we find non-robustness of the means to the prior covariance.



Experiments

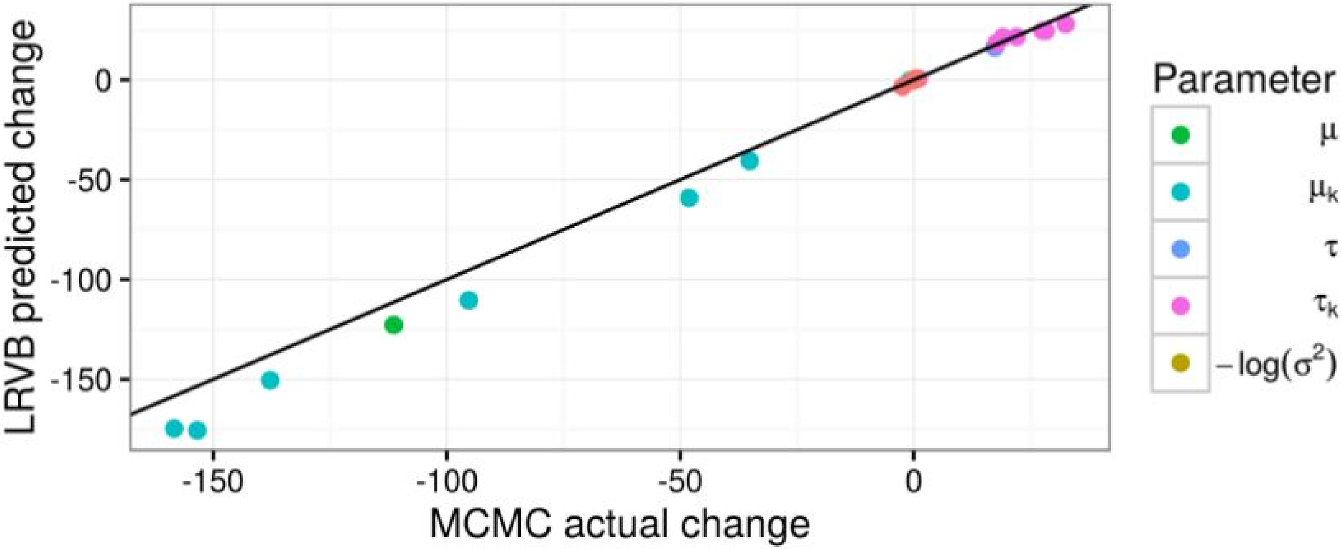
More detailed plots of the sensitivity of τ (average microcredit effectiveness) to different prior parameters.





Experiments

Confirmation of the LRVB sensitivity measurements by manually perturbing and re-running the MCMC.



Practical implication of non-robustness

The posterior for the average effect of microcredit was:

$$\mathbb{E}_q[\tau] = 3.08 \qquad \mathrm{StdDev}_q(\tau) = 1.83.$$

The original prior covariance was $\Lambda = \begin{pmatrix} 0.03 & 0 \\ 0 & 0.02 \end{pmatrix}$.

If instead we had used $\Lambda = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.02 \end{pmatrix}$, then the

95% credible interval would no longer have contained zero.

VB Implementation

- C++ with R frontend
- Uses Stan autodifferentiation libraries (no manual derivative calculations)
- Code and simulated data available at

`https://github.com/rgiordan/MicrocreditLRVB`

Fast robustness quantification with variational Bayes. Ryan Giordano, Tamara Broderick, Rachael Meager, Jonathan Huggins, Michael Jordan (2016)

Conclusion

- Hierarchical models are valuable tools in the social sciences, but it is important to check robustness.
- “Local sensitivity” measures how robust posterior expectations are to local changes in the prior.
- We provide fast, accurate local robustness measurements in hierarchical models.
- Forthcoming work: generic non-parametric perturbations, more experiments.